

DOCUMENT RESUME

ED 219 409

TM 820 420

AUTHOR Skinner, Harvey A.; Lei, Hau
 TITLE Manual for Modal Profile Analysis.
 PUB DATE Nov 80
 NOTE 33p.
 EDRS PRICE MF01/PC02 Plus Postage.
 DESCRIPTORS *Classification; *Computer Programs; Program
 Descriptions; Research Design; *Statistical
 Analysis
 IDENTIFIERS FORTRAN Programing Language; *Modal Profile
 Analysis

ABSTRACT

Modal Profile Analysis (MPA) is a Fortran IV program for the identification and cross validation of relatively homogenous subgroups within a dimensional or ordination space. This classification approach is designed for use with several samples in a multiple replication design. The program consists of three stages: (1) identification of preliminary types within each of m samples, (2) replication of these types across samples, and (3) generalizability of the typology with new samples. The computations are based on a sequential application of the singular value decomposition. An important feature of this program is that it allows one to differentiate the independent contribution of elevation, scatter, and shape parameters in defining profile similarity. MPA is based on a dimensional (ordination) approach to classification, although one may identify discrete clusters within the low dimensional space. This user's manual contains four sections: Introduction; Job Set-up; An Example; and Implementation Instructions. (Author/PN)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

MANUAL FOR MODAL PROFILE ANALYSIS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

H. A. Skinner

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

MANUAL FOR MODAL PROFILE ANALYSIS

HARVEY A. SKINNER PH.D. AND HAU LEI M.Sc.

ADDICTION RESEARCH FOUNDATION

33 RUSSELL STREET

TORONTO, CANADA

M5S 2S1

November, 1980

© 1980 by Harvey A. Skinner and Hau Lei

ABSTRACT

A FORTRAN program is described for the identification and cross validation of relatively homogeneous subgroups within a dimensional or ordination space. This classification approach is designed for use with several samples in a replication design. The program consists of three stages: (1) identification of preliminary types within each of m samples, (2) replication of these types across samples, and (3) generalizability of the typology with new samples. The computations are based on a sequential application of the singular value decomposition. An important feature of this program is that it allows one to differentiate the independent contribution of elevation, scatter, and shape parameters in defining profile similarity. Modal Profile Analysis is based on a dimensional (ordination) approach to classification, although one may identify discrete clusters within the low dimensional space. This user's manual contains four sections: I. Introduction, II. Job Set-up, III. An Example, and IV. Implementation Instructions.

I. INTRODUCTION

Since the pioneering work by Sokal and Sneath (1963), there has been an explosive growth in the development of quantitative techniques for classification research or "numerical taxonomy". Although a plethora of classification techniques exists, there is a need for research strategies that systematically address the internal and external validation of identified types (Cormack, 1971). The purpose of this paper is to describe a computer program, termed Modal Profile Analysis, that provides an unified strategy for classification research (Skinner & Lei, 1980). Emphasis is placed upon the (1) identification of preliminary types within several samples, (2) replication of these types across samples, and (3) generalizability of the final typology to new groups of subjects. Based on a dimensional approach to classification, this model is directed to large samples in a multiple replication design. Modal Profile Analysis allows one to differentiate the independent contribution of profile elevation, scatter and shape parameters (Cronbach and Glaser, 1953; Skinner, 1978).

A detailed development of the rationale and computational strategy underlying Modal Profile Analysis is given by Skinner (1977, 1978, 1979a). This model has been used to derive typologies in several contexts. For example, Skinner, Jackson and Hoffmann (1974) identified types among three samples of alcoholic patients; Skinner, Reed and Jackson (1976) examined the generalizability of these types with psychiatric and prison samples; Burger and Cross (1979) derived personality types as measured by the California Psychological Inventory (Gough, 1969); Skinner and Jackson (1977) developed personality types using the Personality Research Form; and finally, Skinner and Jackson (1978) and Skinner (1979b) developed a model of psychopathology based on three MMPI superordinate types.

Geometrically, the model seeks to represent relationships among subjects in a low dimensional space. Least-squares estimates of parameters are obtained using the singular value or Eckart-Young decomposition (Stewart, 1973). In the final solution one may interpret the dimensions themselves (eg. neuroticism) and/or search for clusters or modes within the dimensional space (Skinner, 1979b).

The Model

For a data matrix X giving the observations of N subjects on k attribute measures, the N subjects may be partitioned by row into m samples of size n_j , where X_j represents the partitioned data matrix for sample j . The fundamental equation of Modal Profile Analysis states

$$X_j = f \left\{ M_j, S_j, (A_j P' + E_j) \right\}$$

That is, relationships among subject profiles (row vectors) of sample j may be expressed as a function of (1) elevation parameters M_j (subject's mean over the k attribute measures), (2) scatter parameters S_j (subject's standard deviation), and (3) a matrix of shape parameters A_j that describe the degree to which a subject resembles the "ideal types" or modal profiles contained in P . Both M_j and S_j are column vectors ($n_j \times 1$), A_j is of order ($n_j \times i$) and P is of order ($k \times i$), where i denotes the number of types in the classification. E_j is a ($n_j \times k$) residual matrix that provides a measure of how well the model fits sample j .

With an instrument such as the Minnesota Multiphasic Personality Inventory (Hathaway and McKinley, 1967), the elevation parameters M_j may be interpreted as a global index of symptom severity or maladjustment (Skinner and Jackson, 1978). The shape parameters in A_j provide coordinates for a dimensional plot. Subjects that have a high positive weight (cluster at the positive pole of the axis) would have a MMPI profile that corresponds closely to the theoretical modal

profile contained in the appropriate column of P. Skinner and Jackson (1978) found a three dimensional solution on the MMPI with axes corresponding to Neurotic, Psychotic and Sociopathic types.

Computational Strategy

The program consists of three stages. First, preliminary types are derived separately within each of m samples. Essentially, this stage involves a singular value decomposition (Stewart, 1973) of the data matrix X_j for sample j that has been scaled to remove the elevation and scatter parameters. This analysis is repeated with each of the m samples under study. In the second stage, the preliminary types from the m samples are compared to determine which types replicate. The stage involves a singular value decomposition of the supermatrix of preliminary types from the m samples. The output from this second stage is a set of final types (matrix P). In the third stage, the final typology may be used to classify new samples. Also, the third stage allows one to estimate classification hit-rates due to chance (Skinner 1979a). This chance baseline may be compared with actual hit-rates to give an estimate of the degree to which the classification solution represents an improvement over chance. The output from Stage III corresponds to the model described above.

II. JOB SET-UP

INPUT

The program is actually a collection of three subprograms each of which corresponds to one phase (or stage) of the Modal Profile Analysis (MPA). The computer job, therefore, should be set up for the one appropriate stage only. Moreover, the program assumes that the input data matrix is "clean". Consequently, all necessary data transformations and missing data corrections should be done before entering MPA. The structure of the input data matrix for each stage is defined in the job set-up below.

A. Selection Card is always the first card (following the JCL cards):

Column 1 IStage = 1 for Stage I: Within-Sample Analysis
 = 2 for Stage II: Multiprofile-Multisample Analysis
 = 3 for Stage III: Classification Analysis.

B. Proceed to the appropriate stage and prepare input accordingly:

STAGE I: WITHIN-SAMPLE ANALYSIS(a) Title Card

Column 1 - 80 TITLE

= A title for the job, up to 80 columns, can be provided here. It will be printed at the beginning of the output.

(b) Number of Samples Card

Column 1-3 NSAMP

= An integer (all the numbers are right-justified unless specified otherwise) corresponds to the number of samples of the population to be analyzed.

(c) Sample Label Card

Column 1-80 TITLE

= A label for the sample data matrix, up to 80 columns, can be specified here. It will be printed at the beginning of the output related to the given sample.

(d) Parameter Card

Column 1-3 NS

= Number of subjects for the given sample.

4-6 NC

= Number of variables for each subject.

7-9 NF

= Number of preliminary profiles hypothesized
= 999 if number of profiles is to be determined by the program. The program uses a criterion similar to eigenvalues greater-than-one to decide the number of preliminary profiles for each sample (See Page 15).

12 KP

= 1 If input data matrix is to be printed.

13-15 EPS

= The absolute value of minimum loading on a profile that a subject should have in order to be classified into the profile. If it is uncertain, use 0.5 (or 500; since there is no decimal point in the field the program will take them as decimals).

18 KS

= 1 If the variables are to be standardized for the input data.

19-21 NFIL

= Logical file unit of input data, e.g. 5 for Card reader (See Page 22).

22-24 MPPUN

= 0 The preliminary profiles are printed only.

= 7 The preliminary profiles are not only printed but also punched on cards.

(d) Parameter Card (cont'd)

= XX The preliminary profiles are not only printed but also recorded on logical file unit xx, e.g., 20 (See Page 23).

(e) Format Cards

Column 1-80

FORTRAN format statement for reading the input cards. {e.g. (3X,12F6.3)} Two cards must be present but the second one can be left blank.

(f) Input Data Cards

Column 1-80

Data are read in subject by subject. Suppose each of the NS subjects has NC variables recorded on r cards (or records). Then there are NS x r cards (records) in this section of input and every r consecutive cards (records) contain the NC variables for a given subject.

(g) Variable Label Cards:

Column 1-21

A label up to 21 columns can be provided for each of the NC variables. Each label should be right-justified, that is, a d-column long label should be started on (22 - d)th column. There are NC cards for this section.

(h) Repeat (c) to (f) inclusive for each subsequent sample.

STAGE II: MULTIPROFILE-MULTISAMPLE ANALYSIS(a) Title Card

Column 1-80 TITLE

Same as in STAGE I

(b) Parameter Card

Column 1-3 MP

= Number of input preliminary profiles (They are usually from the output of STAGE I).

4-6 NC

= Number of variables in each profile.

7-9 NF

= Number of modal profiles hypothesized.

12 KP

= 1 if input data matrix is to be printed.

13-15 EPS

= The absolute value of minimum loading on a profile that a variable should have in order to be a marker in the rotation (See Page 16). If it is uncertain, use 0.5 (or 500; since the program will read them as decimals when no decimal point is encountered in the field.)

16-18 NFIL

= Logical file unit of input profiles, e.g., 5 for card reader (See Page 22).

19-21 MPPUN

= 0 The modal profiles are printed only.

= 7 The modal profiles are not only printed but also punched on cards.

= XX The modal profiles are not only printed but also recorded on logical file unit XX, eg., 20. (See Page 23).

(c) Format Cards

Column 1-80

FORTRAN format statement for reading the input profiles. Two cards must be present. If only one card is needed for the format, it must be followed by a blank card.

(d) Input Profile Cards

Column 1-80

Data are read in profile by profile. Suppose each of the MP profiles has NC variables recorded on r cards (or records). Then there are MP x r cards (records) in this part of input and every r consecutive cards (records) contain the NC variables for a given profile.

(e) Variable Label Cards

Column 1-21

A label up to 21 columns can be provided for each of the NC variables. Each label should be right-justified. {Also see STAGE I: (g).}

STAGE III: CLASSIFICATION ANALYSIS(a) Title Card

Column 1-80 TITLE

= Same as in STAGE I.

(b) Parameter Card

Column 1-3 NV

= Number of variables in the profile.

4-6 MP

= Number of modal profiles (usually from output of STAGE II).

7-9 NG

= Number of samples to be classified.

12 IZ

= 1 If Fisher's R transformation for shape parameters is requested. (See Page 19).

13-15 IP

= 0 If each subject's elevation, scatter, and shape parameters are to be printed only.

= 7 In addition to the printed output, punched cards are requested for each subject.

= XX In addition to the printed output, the parameters of each subject are recorded on logical file unit XX, e.g. 20 (See Page 23).

18 IPERT

= 1 If Modal profiles are to be perturbed (See Page 19).

19-23 IRAN

= Any five-digit number to initiate the random number generation function for perturbation. It can be left blank if IPERT above is not one.

24-26 EPS

= Same as in STAGE I. The absolute value of minimum loading on a modal profile that a subject should have in order to be classified into that dimension (modal profile).

27-29 INMP

= Logical file unit of the input modal profiles, e.g., 5 for card reader (See Page 22).

30-32 NFIL

= Logical file unit of input data cards, e.g., 5 for card reader (See Page 22).

(c) Sample Sizes Card

Column 1-3 NS(1)

= Number of subjects in the first sample.

4-6 NS(2)

= Number of subjects in the second sample.

58 -60 NS(20)

= Number of subjects in the 20th sample.

(d) Standardization Card

Column 1-3 ISS(1)

= 1 If the variables are to be standardized in the first sample.

4-6 ISS(2)

= 1 If the variables are to be standardized in the second sample.

⋮

58-60 ISS(20)

= 1 If the variables are to be standardized in 20th sample.

(e) Format for the Modal-Profile Cards

Column 1-80 Two cards must be present {See STAGE II (c)}

(f) Input Modal-Profile Cards

Column 1-80 See STAGE II (d).

(g) Sample Label Card

Column 1-80 See STAGE I (c).

(h) Format for the Input Data Cards

Column 1-80 Two cards must be present {See STAGE I (e)}.

(i) Input Data Cards

Column 1-80 See STAGE I (f)

(j) Repeat (g) to (i) inclusive for each subsequent sample.

INPUT-OUTPUT DEVICES

The program provides flexibilities for input-output by allowing the user to specify the logical file unit for the data file. When the logical file unit is other than 5, 6, or 7 (usually for card reader, line printer, or card punch, respectively), the user is required to prepare corresponding JCL cards for the given data file. For example, if the user has input data with the DSN=USER.INPUT.DATA stored on disk and he specifies the logical file unit of the input as 20 (e.g. NFIL IN STAGE I(d)), a JCL card similar to the following should be prepared:

```
//GO.FT20F001 DD DSN=USER.INPUT.DATA,DISP=OLD
```

If the user would like the output of preliminary profiles recorded on disk with DSN=USER.OUTPUT.DATA and specifies the logical file unit of the output as 21 (MPPUN in STAGE I(d)), then the following JCL card should be provided:

```
//GO.FT21F001 DD DSN=USER.OUTPUT.DATA,
// UNIT=USER,SPACE=(TRK,(10,5),RLSE),DISP=(NEW,CATLG),
// DCB=(RECFM=FM,LRECL=80,BLKSIZE=12960)
```

These examples are for illustration purposes only. The JCL shown is used at the University of Toronto Computing Center, and may need modification at other installation. Consult your local computer personnel regarding the specifics.

PROGRAM LIMITATIONS

Version I of MPA has been programmed for a maximum of :

- 40 variables in each profile,
- 200 subjects within each sample.

Samples larger than 200 subjects can be either randomly broken down into samples with 200 subjects and less, or handled by the program after re-dimensioning the COMMON block. The modification of MPA to handle larger or smaller data sets is discussed under the Implementation Instruction section.

PRINTED OUTPUT

Output produced by the program is listed as follows:

- (1) Input Data: The program will first echo the input parameters such as the number of subjects, variables, and factors (or profiles). The input data matrix can be also printed upon request. It will be a subject by Variable matrix for Stage I and Profile by Variable matrix for Stage II. The input modal profiles and their correlation matrix are always printed in the output of Stage III.
- (2) Mean and Standard Deviation (S.D.) for Each Variable: This will be printed only if standardization by variable is requested.
- (3) Mean and S.D. for Each Subject: They are the elevation and scatter parameters of the subject. This is printed only in the output of Stage I.
- (4) Principal Axis Factor Matrix: This is a full-rank unrotated factor loading matrix from principal component analysis. For Stage I, the solution is based on a Subject by Subject correlation matrix and the maximum number of factors is equal to the number of variables. For Stage II, the solution is based on a Profile by Profile correlation matrix and the maximum number of factors is equal to the total number of preliminary profiles.
- (5) Eigenvalues for Each Factor and Factor Scores for Each Variable.
- (6) Preliminary Varimax Solution: This is a reduced-rank varimax-rotated factor loading matrix from principal component analysis. The number of factors rotated (i.e. number of profiles hypothesized) is either specified by the user or determined by the program. Successive variances and checks of communalities are also printed.

(7) Hypothesis Matrix and Transformation Matrix for Univocal Varimax Rotation:

The hypothesis matrix is obtained by placing one for each variable on the factor on which the variable has the highest loading and zeros elsewhere. The univocal varimax rotation then uses the hypothesis matrix as a target to transform the preliminary varimax solution (Jackson & Skinner, 1975).

(8) Rotated Matrix: This is an univocal-varimax-rotated factor loading matrix for the Subject by Subject (Stage I) or Profile by Profile (Stage II) correlation matrix.

(9) Rotated Factor Scores: This is the factor score matrix corresponding to the univocal varimax-rotated factor loading matrix above. It provides scores on the variables for each modal profile. The scores are standardized (mean = 0, S.D. = 1) within each profile. A T-score version of the factor scores is also printed along with the variable labels at the end of the output. This is essentially a rescaling of the original factor scores so that they have a mean of 50 and S.D. of 10 within each profile.

(10) Classification Analysis: Both Stage I and III (not Stage II) generate output of the classification analysis. It includes the subject's highest loading, profile (or type) assigned, and communality. It also reports the number of subjects classified in each profile (positive vs. negative poles), classification efficiency, and the total percentage of variance explained. For Stage III, the classification may be based on a perturbed profile matrix. If the perturbation is requested, the perturbed profile matrix will be printed. The elevation, scatter, and shape parameters are listed for each subject in the printout of Stage III. If Fisher's R to Z transformation is requested, the

(10) Classification Analysis (cont'd)

transformed shape parameters for each subject will be printed next. Moreover, punched output of elevation, scatter, and shape parameters can be requested also. Note if Fisher's transformation is requested in the same run, the transformed shape parameters will be punched instead of the original shape parameters.

III AN EXAMPLE

The complete output for the following sample problem is provided with the source copy of MPA. One may use this sample problem for testing MPA on your computer system. The data are taken from 88 archetypal psychiatric patients that were used by Mezzich (1978) in an evaluation of various clustering methods. The N=88 archetypal patients were developed by experienced psychiatrists to represent four diagnostic categories: 1) Manic-Depressive Depressed (MDD), 2) Manic-Depressive Manic (MDM), 3) Simple Schizophrenic (SS), and 4) Paranoid Schizophrenic (PS). Each archetypal patient is characterized by scores on 17 scales, such as, Anxiety, Hostility, and Hallucinatory Behavior. Following Mezzich (1978), the archetypal patients were divided into two samples A and B of 44 patients each. Within each sample, there were 11 examples of each diagnostic category (MDD = MDM = SS = PS = 11).

Stage I. Within-Sample Analysis

In general the output from MPA follows the sequence of equations given in Skinner (1979a). Initially, the program prints the various parameters specified and (optionally) the input data matrix. If the variables are not of a comparable scale (eg. standard scores), then the measures may be converted to z scores (mean=0, variance=1) by specifying KS=1 in column 18 of the Parameter Card.

Next, the mean and standard deviation are printed for each subject, which are respectively the elevation and scatter parameters of equation 19 in Skinner (1979a). For example, with the sample A data, the elevation parameters range from 1.353 for Subject 26 to 4.235 for Subject 44. Then, results from the singular-value decomposition of the data matrix are printed. The Principal Axes Factor Matrix corresponds to equation 2 in Skinner (1979a), and describes the weightings of subjects on factors in the person space (Q-technique). The

Eigenvalues (r^2) depict the relative importance of each factor. In this example, there is a large first eigenvalue (21.4127), followed by three eigenvalues of lesser magnitude (7.8499, 4.6980, 3.3016, respectively), while the remaining dimensions are of minor significance. The first four dimensions account for 84.68% of the variance. Finally, the Principal Axes Factor Scores are printed (W_j from equation 1 in Skinner, 1979a). This matrix gives the projection of variables upon factors in the variable space.

The investigator may either specify the number of factors to retain as NF in Column 9 of the Parameter Card, or the NF=999 option will use an internal criterion to estimate the number of dimensions that should be retained for interpretation. Basically, this criterion is based on the number of eigenvalues that exceed 1.0 if one had conducted a principal components analysis of the variables (R-technique). Computationally, this is done by rescaling the Eigenvalues (r^2) by NC (variables)/NS (subjects). This scaling changes the Eigenvalues which sum to the total number of subjects (NS) to rescaled Eigenvalues that sum to the total number of variables (NC). Then, the eigenvalues greater-than-one criterion is applied to the rescaled Eigenvalues. In practice, I have found this criterion to give a reasonably good first estimate of the number of factors to interpret. Subsequently, I examine the distribution of eigenvalues (scree test) to look for marked breaks. Also, one should examine the interpretability of the rotated solution as another criterion in deciding upon the final number of factors. In the present example, the NF=999 option resulted in the retention of 4 factors.

The next matrices printed describe the results from a Univocal Varimax rotation of the subject factors (Jackson & Skinner, 1975). First, the Varimax solution is printed, followed by the Hypothesis Matrix where a subject has been targeted for a single factor if the initial Varimax weighting exceeds |.50|. The Rotated Matrix represents an attempt to have each subject maximally

load on only one factor. In the example, the first 11 subjects (Manic-Depressive Depressed) all weight primarily on factor 3, the next 11 Manic-Depressive Manic prototypes are mainly loaded on factor 4 (except for Subjects 20 and 21), the 11 Simple Schizophrenic prototypes all load on factor 2, and the final 11 Paranoid Schizophrenic prototypes are mainly located on factor 1.

The subsequent output focuses upon an examination of each factor (Type) to determine which subjects have a salient weighting on it. The minimum loading that a subject should have to be considered salient is specified by the EPS parameter in Columns 13-15 of the Parameter Card. Generally, I have used an absolute value of .50. In the current example, 14 subjects were salient on factor (Type) 1, 11 subjects on Type 2, 11 subjects on Type 3, and 8 subjects on Type 4. Three subjects (13,20,21) were not grouped with their like-named prototypes. The Classification Efficiency provides one index of how well the model fits the data. A reasonably high percentage of subjects must be salient on the factors if one is to achieve a meaningful interpretation of the data. The Total Percentage of Variance Explained is a second measure of how well the principal component model fits the data.

Corresponding to factors in the person space (Q-technique), the next matrix printed is the Rotated Factor Scores which gives the characteristic profile of variables that corresponds to the subset of subjects that are salient on a given factor (equation 4 from Skinner, 1979a). This matrix is used for input to Stage II. Finally, this matrix is scaled to resemble T-scores* and printed in integer form as the modal Profiles. In our example, the first Profile is characterized by high scores on Suspiciousness, Unusual Thought Content, and Uncooperativeness, while there is an absence or low scores on Motor Retardation, Guilt Feelings and Blunted Affect. This description corresponds favourably and with assumed characteristics of Paranoid

* mean = 50, standard deviation = 10

Schizophrenic patients (subjects 34 to 44).

The remaining output presents a parallel analysis of the 44 archetypal patients in Sample B. In this case, 4 factors were also retained and only one subject (18) was not correctly placed with the like-named prototypes.

Stage II. Multiprofile-Multisample Analysis

Since 4 factors were retained in both Sample A and Sample B, the second stage involves an analysis of the congruence between modal profiles from the two samples.

Following a listing of the input modal profiles, they are intercorrelated. Each profile should have a mean = 0.00 and variance = 1.00. Otherwise, an error must have occurred in carrying forth the modal profiles from Stage I. In this event, check to make sure that the output from Stage I has been correctly read into Stage II. Since an orthogonal rotation (Jackson & Skinner, 1975) was conducted in Stage I, the within-sample profiles must be statistically independent, that is, they must not be correlated. Hence, within set partitions of the correlation matrix must yield an identity matrix (1.0 in the diagonal, 0.0 elsewhere). If not, then some error must have been made in carrying forth the modal profiles from Stage I.

**Note **

If the means are not equal to 0.0, variances not equal to 1.0 and within-sample correlations not equal to 0.0, then DO NOT interpret the rest of the output. Determine where the error occurred in linking the Stage I modal profiles to Stage II.

Next, the singular-value decomposition is printed which corresponds to equations 6 and 7 from Skinner (1979a). The Principal Axes Factor Matrix describes the projection of the within-sample modal profiles or principal axes of the person space, the Eigenvalues depict the relative importance of each

factor, and the Principal Axes Factor Scores describe the variable space. As discussed by Skinner (1979a), with two samples the (Eigenvalues - 1.0) may be interpreted as canonical correlations for "matching" factors across the two samples. In this example, the canonical correlations are respectively .9901; .9653, .9391, and .6498. This distribution suggests that certainly three and likely four factors should be retained for interpretation. The Rotated Matrix indicates that one modal profile from Sample A clearly corresponds to one modal profile in the second sample. For instance, the second modal profile from Sample A (#2 weighting = .976) and the second modal profile from Sample B (#6 weighting = .976) define the first inter-battery factor of equation 9 (Skinner, 1979a).

Finally, the corresponding modal profiles are printed (equation 10; Skinner, 1979a). For example, the first profile is marked by elevations on Blunted Affect, Emotional Withdrawal, and Conceptual Disorganization in contrast to an absence of Hallucinatory Behaviour. Since the profiles are given in T-score format, scores above 60 are clearly present whereas scores below 40 tend to be absent. An individual who had a high positive weighting on this profile would have a profile of variables that closely corresponded to the pattern of high and low scales. In this context, a positive weighting would reflect a confused, emotionally withdrawn individual. Conversely, a high negative weighting would indicate a person who exhibited hallucinatory behavior but a relative absence of the emotional withdrawal syndrome.

Stage III. Classification Analysis

To illustrate the Classification Analysis, the modal profiles from Stage II were used to classify the original subjects in Samples A and B. The analysis corresponds to equations 16, 17 and 18 in Skinner (1979a). First, subjects are correlated with the four modal profiles, and a classification summary is

printed. In Sample A, only one subject (#21) was not grouped with a like named archetypal patient. The model classifies 100% and accounts for 82.74% total variance in Sample A. Finally, the Elevation, Scatter and Shape parameters are printed which correspond to equation 19 (Skinner, 1979a). If these data are to be used subsequently as predictors (or correlated with) other variables, then one should consider specifying the Fisher's r to z transformation of the shape parameters (IZ in column 12 of the Parameter Card).

Also, Stage III may be used to estimate a chance baseline for fitting the model. Here, the input modal profiles would be randomly perturbed by specifying IPERT = 1 in Column 18 of the Parameter Card, as well as by specifying a five-digit random number to initiate the random number generating function (IRAN in Column 19-23). As discussed by Skinner (1979a) on pages 331-332, one may repeat the perturbation strategy with a "fresh" random number in IRAN for each run to yield a sampling distribution of the chance baseline.

IV. IMPLEMENTATION INSTRUCTIONS

MPA is written in ANSI standard FORTRAN IV. It has been implemented on IBM 360-370 systems. It makes extensive use of a matrix-operation package, MATPAK, developed by Jackson, Skinner & Strasburger (1977). The main program of MPA calls 29 subroutines among which 20 are from MATPAK and one from IBM. The names of all the subroutines are listed in Table

ALTERING PROGRAM LIMITATIONS

The program limitations for MPA may be modified by changing the COMMON statement in the program source file. All the corresponding COMMON statements in the main program and subroutines should be replaced. In the present version, the COMMON block is specified as:

```
COMMON  A(3, 200, 40), C(400), NORD (3,2), IX, NMP
```

where 200 is the present limitation for the number of subjects within a sample (i.e. NS),

40 is the present limitation for the number of variables within a profile (i.e. NC),

400 is two times the number of subjects within a sample (i.e. $2 \times NS$).

To update the COMMON statement, the user simply substitutes the desired NS, NC, and $2 \times NS$ in the corresponding locations. For example, if you would like to analyze a data matrix in which each of the 300 subjects is measured on 30 variables, the appropriate dimensions would be:

```
COMMON  A(3, 300, 30), C(600), NORD (3, 2), IX, NMP
```

Note the maximum dimensions you can specify is ultimately determined by the memory capacity of your computer installation.

CORE MEMORY REQUIREMENTS

The load module of MPA (Version I) has been implemented on the IBM 370/3033/MVS system at the University of Toronto Computer Services(UTCS). The core requirement for executing it varies with the stage of the analysis, specifically, 250K for Stage I, 230K for Stage II, and 210 K for Stage III. The core memory is requested through a keyword parameter, namely, REGION.GO, on the EXEC card. For example, at UTCS:

```
// EXEC  LEGO,REGION.GO=250K
```

is the JCL statement to request 250K for executing MPA.

If the user has re-dimensioned the COMMON block, then corresponding adjustment should be made on the estimate of core requirement. The procedure is illustrated with the following example:

Suppose the new COMMON statement is:

```
COMMON  A(3, 400,40), C(800), NORD(3,2), IX,NMP
```

- (1) Compute the number of elements in the COMMON block, which is:

$$(3 \times 400 \times 40) + 800 + (3 \times 2) + 1 + 1 = 48,808$$

- (2) Multiply the number of elements by 4 (bytes):

$$48,808 \times 4 = 195,232 \text{ (bytes)}$$

- (3) Divide the number of bytes by 1024 bytes to get K-bytes:

$$195,232 / 1024 = 190.7 \text{ (K-bytes)}$$

- (4) Find the difference between the new COMMON area and the original COMMON area

$$\text{(which is 96 K): } 191 \text{ K} - 96 \text{ K} = 95 \text{ K}$$

- (5) Add the difference to the original core requirement (e.g. 250K)

$$250 \text{ K} + 95 \text{ K} = 345 \text{ K}$$

- (6) Request the updated core memory by:

```
// EXEC  LEGO,REGION.GO=345K
```

You may use 1000, instead of 1024 at Step (3) above, to make the calculation easier. Consequently, you will have a more generous estimate for the core requirement (e.g. 350K in this case). You may not be able to have this convenience, however, when the core

requirement is on the border of limitation of the specific computer facility.

EXECUTE MPA FROM A LOAD MODULE

For the user of UTCS, MPA is executed from a load module and it can be accessed by the following JCL (Job Control Language) statements:

```
//      JOB      (UTCS  JOB CARDS)
//      EXEC  LEGO,REGION.GO=250K
//LE.SYSLIN DD*
        INCLUDE LIB(MPA)
/*
//LIB  DD  DSN=USER.ARF.HAU.LODLIB,DISP=SHR
//GO.FT37F001 DD  DSN=&&TEMP,UNIT=SYSDA,SPACE=(CYL,(1,1)),
//  DCB=(LRECL=32000,BLKSIZE=13030,RECFM=VBS)
//GO.SYSIN DD *

        MPA PARAMETER CARDS AND DATA CARDS
/*
//
```

The above JCL setup assumes that MPA parameters and data are submitted together with JCL via the same input device (i.e. card reader or terminal). If they are separated from JCL statements, appropriate modification should be made according to the specific situation. For example, if both MPA parameters and data are stored as one disk file by the name of USER.MPA.DATA, the JCL statement of

```
//GO.SYSIN DD *
```

is replaced by: //GO.SYSIN DD DSN=USER.MPA.DATA,DISP=OLD

Then the statement is immediately followed by /* and // cards.

The MPA parameters and data can also be stored as one tape file. An example is shown in the next section (EXECUTE MPA FROM A TAPE).

Furthermore, data (observations or preliminary profiles) can be separated from

MPA parameters and read in from any input device (e.g. tape or disk). The corresponding JCL statements, however, should be prepared and inserted in front of the GO.SYSIN card. For example, if the data file by the name of MY.INPUT.DATA is stored on tape and Columns 20-21 on the parameter card of Stage I is specified as 20, then statements similar to the following should be inserted:

```
//GO.FT20F001 DD DSN=MY.INPUT.DATA,DISP=OLD,
// DCB=(DEN=3),UNIT=3400-3,
// LABEL=(2,SL),VOL=SER=ABCDEF
```

The above tape is assumed to be 1600 BPI (i.e. DEN=3), 9 track, with an internal volume name of "ABCDEF". It is standard labelled (i.e. SL), and the data is the second file on tape. In addition, a setup card similar to the following should follow the JOB card:

```
/*SETUP UNIT=3400-3,VOL=(INT=ABCDEF,EXT=F1234)
```

This requests a tape drive suitable for the tape (UNIT=3400-3), and specifies the slot (F1234) where the tape is stored.

Similarly, the optional output of MPA can be directed to various devices such as disk and tape. For example, the following JCL statements are for writing output file on disk:

```
//GO.FT21F001 DD DSN=USER.OUTPUT.DATA,
// UNIT=USER,SPACE=(TRK,(10,5),RLSE),DISP=(NEW,CATLG),
// DCB=(RECFM=FB,LRECL=80,BLKSIZE=12960)
```

The number 21 (in GO.FT21) should be specified in the appropriate columns of MPA parameter card (e.g. columns 23-24 on the parameter card of Stage I.) If you are not familiar with the JCL conventions, consult computer personnel before attempting any of the above options.

EXECUTE MPA FROM A TAPE

For users who are planning to modify the dimensions of the program or who do not have access to UTCS, a tape copy of MPA will be provided by the authors at cost upon request. The sample problem corresponding to the EXAMPLE section of this manual is

also included on the tape so that the user may test MPA on his system.

Using the sample problem on the same tape (MPA parameters and data together as one file), the following JCL executes the Stage I of MPA from the tape provided by the authors. The JCL is based on UTCS's specifications. Users of other installations are advised to consult their local computer personnel for necessary changes.

```
//      JOB              (UTCS  JOB  CARDS)
/*SETUP  UNIT=3400-5,VOL=(EXT=F1234,INT=MPATAP)
//      EXEC  FORTLEGO,REGION=250K
//FORT.SYSIN DD  DSN=ARF.SKINNER.MPA.SOURCE,DISP=(OLD,KEEP),
//      DCB=(LRECL=80,BLKSIZE=12960,RECFM=FB,DEN=4)
//      UNIT=3400-5,LABEL=(1,SL,,IN),VOL=SER=MPATAP
//GO.FT37F001 DD DSN=&&TEMP,UNIT=SYSDA,SPACE=(CYL,(1,1)),
//      DCB=(RECFM=VBS,LRECL=32000,BLKSIZE=13030)
//GO.SYSIN DD  DSN=MPA.STAGE.ONE.DATA,DISP=(OLD,KEEP),
//      DCB=(LRECL=80,BLKSIZE=12960,RECFM=FB,DEN=4),
//      UNIT=3400-5,LABEL=(2,SL,,IN),VOL=SER=MPATAP
```

Note: F1234 should be substituted with the real slot number that the tape is allocated at the UTCS (or your installation).

For the Stage II analysis of the sample problem replace the GO.SYSIN cards with:

```
//GO.SYSIN DD  DSN=MPA.STAGE.TWO.DATA,DISP=(OLD,KEEP),
//      DCB=(LRECL=80,BLKSIZE=12960,RECFM=FB,DEN=4),
//      UNIT=3400-5,LABEL=(3,SL,,IN),VOL=SER=MPATAP
```

Similarly, for the Stage III analysis of the sample problem, change the DSNAME to DSN=MPA.STAGE.THREE.DATA and LABEL to LABEL= (4,SL,,IN). To analyze your own data, simply change the GO.SYSIN card to suit your specifications (See also the preceeding section on how to read data separately.)

Depending on the frequency of the use of the program, other options are available in implementing your copy of MPA. One option is to copy the program on tape onto a disk and then do the redimension (if necessary) on the disk file. The

updated version of MPA can be copied back to the tape and run from there if your use of program is infrequent. If the use is frequent, then the disk program file may be compiled to create an object or load module on disk to avoid recompiling each time you use the program.

One way to create a load module disk file is to run the following job (Again, the JCL follows the specifications at UTCS):

```
//      JOB          (UTCS JOB CARDS)
// EXEC  FORTLE, NAME=MPA,LOADLIB='USER.MY.LOADLIB',
//  PARM.LE=(NCAL,LET),LEVEL=H
//FORT.SYSIN DD DSN=USER.MPA.PROGRAM,DISP=OLD
//
```

Note: 1. Use name of your own load library for USER.MY.LOADLIB.

2. Use name of your disk file for USER.MPA.PROGRAM.

3. To create load module directly from the tape, use:

```
//FORT.SYSIN DD DSN=ARF.SKINNER.MPA.SOURCE,DISP=OLD,
// DCB=(LRECL=80,BLKSIZE=12960,RECFM=FB,DEN=4)
// UNIT=(3400-5),LABEL=(1,SL,,IN),VOL=SER=MPATAP
```

(Also, remember to put SETUP card before EXEC card)

The job will create a non-executable load module named MPA and place it in a load library named USER.MY.LOADLIB which is an existing partitioned data set (if it is not existing, you have to create a dummy file first by allocating space for it.) The non-executable load module is link-edited at execution time to make an executable load module, when it is called by JCL statements similar to what described in the section of "EXECUTE MPA FROM A LOAD MODULE" (Except on LIB card, after DSN=, put whatever name you use for USER.MY.LOADLIB).

Although efforts have been made to detail every possible implementation step so that users can follow the examples like a cookbook, this section does require some working knowledge of JCL. In addition, the JCL setup varies somewhat at different installations. Consequently, JCL-naive users are strongly recommended to have their

local computer consultants go over their job setups before running the program.

REFERENCES

- Burger, G.K., and Cross, D.T. Personality types as measured by the California Psychological Inventory. Journal of Consulting and Clinical Psychology, 1979, 47, 65-71.
- Cormack, R.M. A review of classification. Journal of Royal Statistical Association, 1971, 134, 321-353.
- Cronbach, L.J., and Gleser, G.C. Assessing similarity between profiles. Psychological Bulletin, 1953, 50, 456-473.
- Gough, H.M. California Psychological Inventory. Palo Alto: Consulting Psychologists Press, 1969.
- Hathaway, S.R., and McKinley, J.C. Minnesota Multiphasic Personality Inventory. New York: Psychological Corporation, 1967.
- Jackson, D.N., & Skinner, H.A. Univocal Varimax: Educational and Psychological Measurement, 1975, 35, 663-665.
- Mezzich, J.E. Evaluating clustering methods for psychiatric diagnosis. Biological Psychiatry, 1978, 13, 265-281.
- Skinner, H.A. The eyes that fix you: A model for classification research. Canadian Psychological Review, 1977, 18, 142-151.
- Skinner, H.A. Differentiating the contribution of elevation, scatter and shape in profile similarity. Educational and Psychological Measurement, 1978, 38, 297-308.
- Skinner, H.A. Dimensions and Clusters: A hybrid approach to classification. Applied Psychological Measurement, 1979, 3, 327-341. (a)
- Skinner, H.A. A model of psychopathology based on the MMPI. In C.S. Newmark (Ed.), MMPI: Current clinical and research trends. New York: Praeger, 1979 b.
- Skinner, H.A., and Jackson, D.N. The missing person in personnel classification: A tale of two models. Canadian Journal of Behavioural Sciences, 1977, 9, 146-160.
- Skinner, H.A., and Jackson, D.N. A model of psychopathology based on an integration of MMPI actuarial systems. Journal of Consulting and Clinical Psychology, 1978, 46, 231-238.
- Skinner, H.A., & Lei, H. Modal Profile Analysis: A computer program for classification research. Educational and Psychological Measurement, 1980.
- Skinner, H.A., Jackson, D.N., and Hoffmann, H. Alcoholic personality types: Identification correlates. Journal of Abnormal Psychology, 1974, 83, 658-666.

Skinner, H.A., Reed, P.L., and Jackson, D.N. Toward the objective diagnosis of psychopathology: Generalizability of modal personality profiles. Journal of Consulting and Clinical Psychology, 1976, 44, 111-117.

Sokal, R.R., and Sneath, P.H.A. Principles of numerical taxonomy. San Francisco: Freeman, 1963.

Stewart, G.W. Introduction to matrix computations. New York: Academic Press, 1973.

TABLE 1

Subroutines Used in MPA

I. From MPA

STAGEA	STAGEB	STAGEC	POWER	CLASS
VTITLE	LABEL	ZOUT		

II. From MATPAK

ERROUT	JIFFY	MCOR	MEANS	MGEN
MGET	MMJL	MPRI	MPURG	MPUT
MRED	MTRA	ROTATE	SMUL	SORT
SVD3	UNIVMX	VMX	ZROW	ZSCORE

III. From IBM

RANDU

Plus functions available on IBM 360-370 systems such as ALOG10, SQRT, and ATAN.